

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-327908

(43)Date of publication of application : 30.11.1999

(51)Int.Cl.

G06F 9/44
G06F 13/00

(21)Application number : 10-122090

(71)Applicant : INTERNATL BUSINESS MACH
CORP <IBM>

(22)Date of filing : 01.05.1998

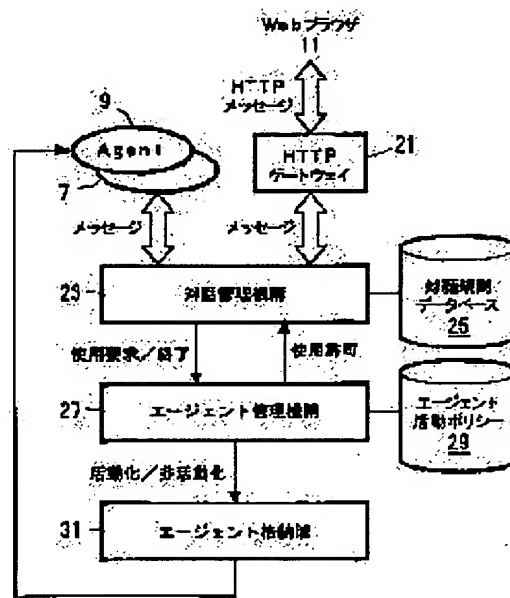
(72)Inventor : YAMAMOTO MANABU
NAKAMURA YUICHI

(54) AGENT ACTIVITY CONTROL METHOD AND COMPUTER

(57)Abstract:

PROBLEM TO BE SOLVED: To make controllable the reduction of the load on an agent server by inspecting a message by referring to an interaction rule and specifying the destination agent of the message as an activation candidate agent.

SOLUTION: An interaction managing mechanism 23 judges whether or not a message is the initial message of the interaction rule. When the message is the initial message of the interaction rule, an interaction management block is generated. Then, an agent activity request is added to a message transmission request. The agent activity request is an activity request to the destination agent of the message and this destination agent becomes an activity candidate agent. Then, this message transmission request is put in a queue. When the message is the initial message of the interaction rule, on the other hand, since the agent is already activated and the interaction management block is generated, an agent managing mechanism 27 is requested to send the message.



LEGAL STATUS

[Date of request for examination] 27.12.1999

[Date of sending the examiner's decision of rejection] 06.11.2001

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection] 2002-01558

[Date of requesting appeal against examiner's decision of rejection] 31.01.2002

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-327908

(43) 公開日 平成11年(1999)11月30日

(51) Int.Cl.⁶

G 0 6 F 9/44
13/00

識別記号

5 5 2
3 5 7

F I

G 0 6 F 9/44
13/00

5 5 2
3 5 7 Z

審査請求 未請求 請求項の数14 O L (全 16 頁)

(21) 出願番号 特願平10-122090

(22) 出願日 平成10年(1998) 5 月 1 日

(71) 出願人 390009531

インターナショナル・ビジネス・マシーンズ・コーポレイション

INTERNATIONAL BUSINESS MACHINES CORPORATION

アメリカ合衆国10504、ニューヨーク州
アーモンク (番地なし)

(72) 発明者 山本 学

神奈川県大和市下鶴間1623番地14 日本アイ・ピー・エム株式会社 東京基礎研究所内

(74) 代理人 弁理士 坂口 博 (外1名)

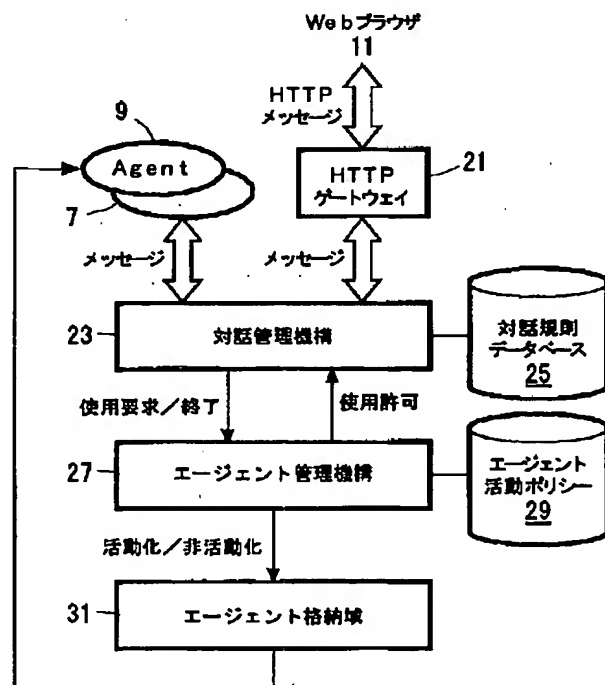
最終頁に続く

(54) 【発明の名称】 エージェント活動制御方法及びコンピュータ

(57) 【要約】 (修正有)

【課題】 エージェント・サーバの負荷を制御する方法及び装置の提供。

【解決手段】 エージェント・サーバは、エージェントの処理が開始する時に二次記憶装置からエージェントを読み込み、エージェントの処理が終了した時に当該エージェントを二次記憶装置に書き出す。これにより、処理を行わないエージェントにより占有されたメモリとスレッドを解放する。また、あるエージェントの処理が終了した後に、別のエージェントの処理を開始することで、同時に活動するエージェントの数を制限する。この処理の開始及び処理の終了を、エージェントの対話の開始及び終了で判断する。メッセージは対話規則において、メッセージの種類から一連の処理の中で交換されるメッセージの流れまでを予め規定しておけば、メッセージ及び当該対話規則を参照することで対話の開始及び終了を検出できる。



【特許請求の範囲】

【請求項 1】 エージェントの活動を制御する方法であって、

使用可能なメッセージを規定する対話規則を参照して、メッセージを検査する検査ステップと、当該検査結果に従い、前記メッセージの宛先エージェントを活動化候補エージェントとして指定する活動化候補エージェント指定ステップと、を含むエージェント活動制御方法。

【請求項 2】 活動中のエージェントの状況を参照して、前記活動化候補エージェントを活動化するステップと、をさらに含む請求項 1 記載のエージェント活動制御方法。

【請求項 3】 前記検査ステップは、使用可能なメッセージ及びエージェント間の対話を開始させるメッセージを定義する対話規則を参照して、メッセージが前記対話を開始させるためのものであるか検査するステップを含む請求項 1 記載のエージェント活動制御方法。

【請求項 4】 前記活動化候補エージェント指定ステップは、前記検査結果により前記メッセージの宛先エージェントは活動化させるべきであると判断された場合には、活動化候補エージェントとして前記宛先エージェントの活動を要求するステップと、前記活動化候補エージェントの活動化の可否を判断するステップとを含む請求項 1 記載のエージェント活動制御方法。

【請求項 5】 前記検査結果により前記メッセージが不正なものであると判断された場合には、前記メッセージの送信元エージェントを非活動化するステップをさらに含む請求項 1 記載のエージェント活動制御方法。

【請求項 6】 第 2 エージェントから送信された複数のメッセージを検査するステップと、前記複数のメッセージが所定の条件を満たしていない場合には、前記第 2 エージェントを非活動化するステップとをさらに含む請求項 1 記載のエージェント活動制御方法。

【請求項 7】 前記検査結果により前記メッセージの送信元エージェントを非活動化させるべきであると判断された場合には、前記送信元エージェントを非活動化候補エージェントに指定するステップ、をさらに含む請求項 1 記載のエージェント活動制御方法。

【請求項 8】 使用可能なメッセージを規定した対話規則を格納した記憶装置と、前記対話規則を参照してメッセージを検査し、当該メッセージの宛先エージェントを活動化させるべきと判断した場合にはエージェント使用要求を出力するメッセージ検査プログラムと、

前記エージェント使用要求を受信して、活動化可能なタイミングで前記宛先エージェントを活動化させるエージェント管理プログラムとを含むコンピュータ。

【請求項 9】 前記対話規則が、エージェント間の対話を開始させるメッセージを定義していることを特徴とする請求項 8 記載のコンピュータ。

【請求項 10】 前記メッセージ検査プログラムは、前記対話規則を参照してメッセージが不正なものであるか判断し、不正なメッセージを検出した場合には、当該メッセージの送信元エージェントの非活動化を要求し、前記エージェント管理プログラムは、当該要求に応答して前記送信元エージェントを非活動化する、請求項 8 記載のコンピュータ。

【請求項 11】 前記メッセージ検査プログラムは、前記対話規則を参照してメッセージを検査し、当該メッセージによりその送信元エージェントに係わる対話が終了すると判断した場合には当該送信元エージェントの非活動化を要求する、請求項 8 記載のコンピュータ。

【請求項 12】 前記記憶装置は、エージェントの可能な活動を規定したエージェント活動ポリシーをさらに記憶しており、前記エージェント管理プログラムは、当該エージェント活動ポリシーを参照して、エージェントの活動化及び非活動化を実施することを特徴とする請求項 8 又は 11 記載のコンピュータ。

【請求項 13】 前記コンピュータはネットワークに接続可能であり、他のコンピュータからのハイパーテキスト・トランスポート・プロトコル (HTTP) メッセージをエージェント用のメッセージに変換する HTTP ゲートウェー・プログラムをさらに有し、前記メッセージ検査プログラムは、当該 HTTP メッセージを変換することにより生成されたメッセージを検査して、前記他のコンピュータに対応するエージェントのエージェント使用要求を出力することを特徴とする請求項 1 記載のコンピュータ。

【請求項 14】 エージェントの活動を制御するプログラムを格納した記憶媒体であって、前記プログラムは、コンピュータに、使用可能なメッセージを規定する対話規則を参照して、メッセージを検査する検査ステップと、当該検査結果により、前記メッセージの宛先エージェントを活動化候補エージェントとして指定する活動化候補エージェント指定ステップと、を実行させる、記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、エージェント技術に関し、より詳しくは、エージェント・サーバにおける

エージェントの活動を制御する方法及び機構に関する。

【 0 0 0 2 】

【従来の技術】 WWW (World Wide Web) 上でのオンライン・ショッピングは、旅行、書籍、CD (Compact Disk) 及び求人情報等幾つかの分野において成功例が報告されており、今後益々発展すると考えられている。このオンライン・ショッピングの形態も、従来までの HTML (Hyper Text Markup Language) ファイル、又は DB (DataBase) と CGI (Common Gateway Interface) のように単に商品を陳列するだけの形態から、複数のサイトを一括検索する形態又は利用者の嗜好を学習して適切な商品を捜し出す形態等に移行しつつある。

【 0 0 0 3 】 このような新たな形態に適しているのがエージェント技術である。例えば、利用者ごとに情報検索エージェントを、出店する会社ごとに情報提供エージェントを生成し、これらのエージェントが対話を行うことにより、情報検索及び商品売買を行うことができる。これらのエージェントは移動エージェントであってもよいが、少なくとも一時的には 1 つのサイト (エージェント・サーバ) 上で一緒に活動する。このサーバ上では利用者の数だけ情報検索エージェントが生成され、それぞれ必要な情報検索を実施するため、エージェント数が大きくなるとサーバにおける負荷は大きくなる。エージェントにより使用されるサーバ資源は主記憶装置とスレッドであるが、情報検索エージェントの場合、情報検索結果を保持すると主記憶装置の占有量が増加する。さらに、エージェント・サーバの利用が増加すると会社の出店も増加して情報提供エージェントも増加するため、さらにサーバの負荷は大きくなる。情報提供エージェントは、通常商品情報をデータベースに保持しており、このデータベース検索は大きな負荷となる。

【 0 0 0 4 】 このようにサーバの負荷制御は大きな問題である。このため従来から、処理を実施していないエージェントをサーバ内の HDD (Hard Disk Drive) などの二次記憶装置に格納することでエージェントの数を減らし、それにより負荷を減少させていた。しかし、エージェント間のメッセージのやり取りを判断して二次記憶装置への格納を実施しているわけではない。また、米国特許第 5 7 0 6 5 1 6 号には、管理者が有効なトークン数を規定し、トークンを取得したエージェントのみが処理を実施することができるような負荷制御について記載がある。しかし、この特許はエージェント間のメッセージのやり取りを判断して負荷制御を行っているわけではない。

【 0 0 0 5 】 ところで、エージェント間でメッセージのやり取りを行い、エージェント間の対話を成立させるためには、当該エージェント間で対話規則が合っていないといけない。そこで、エージェント間の対話規則を規定したタイプ・データベースをエージェント・サーバに設け (例えば、IBM TDB Vol.40 No.12, pp29 Dec. 50

1997 参照のこと)、当該対話規則に従ってメッセージを出力するエージェントを生成すれば、そのエージェント・サーバにおいては、同じくその対話規則に従う他のエージェントとの対話を行うことができる。しかし、プログラムの誤りや故意により、対話規則に従わないメッセージを出力するエージェントや、システムに悪影響を及ぼすように多数のメッセージを出力するようなエージェントがエージェント・サーバにやってくることも考えられる。このような不正なエージェントに対して、従来では対策がなされていなかった。

【 0 0 0 6 】

【発明が解決しようとする課題】 よって本願発明の目的は、エージェント・サーバの負荷を制御することである。

【 0 0 0 7 】 より詳しくは、メッセージを検査することにより、エージェントの活動/非活動を制御する方法を提供することである。

【 0 0 0 8 】 他の目的は、不正なエージェントを検出し、対応処理を実施することである。

【 0 0 0 9 】 より詳しくは、メッセージを検査することにより、不正なメッセージを出力するエージェントを検出し、不正なエージェントを非活動化する方法を提供することである。

【 0 0 1 0 】 なお、エージェントの活動化とはエージェントが処理を実施できる状態にすることをいい、エージェントの非活動化とは再活動化可能な状態で二次記憶装置に格納すること及び場合によっては破棄することである。

【 0 0 1 1 】

【課題を解決するための手段】 エージェント・サーバの負荷を制御するには、(1) 活動していないエージェントを二次記憶装置に格納し、そのエージェントが占有していたメモリとスレッドを解放し、(2) 同時に、活動するエージェントの数を制御する必要がある。エージェント・サーバは、エージェントの処理が開始する時に二次記憶装置からエージェントを読み込み、エージェントの処理が終了した時に当該エージェントを二次記憶装置に書き出す。これにより、処理を行わないエージェントにより占有されたメモリとスレッドを解放する。また、あるエージェントの処理が終了した後に、別のエージェントの処理を開始することで、同時に活動するエージェントの数を制限する。本発明では、この処理の開始及び処理の終了を、エージェントの対話の開始及び終了で判断する。先に述べたように、メッセージは予め規定された対話規則に従う。よって、この対話規則において、メッセージの種類から一連の処理の中で交換されるメッセージの流れまで規定しておけば、メッセージ及び当該対話規則を参照することにより、対話の開始及び終了を検出できる。これにより、処理の開始及び終了を検出できる。

5

【0012】不正なエージェントであるかどうかは、エージェントが送信するメッセージの内容と送信頻度を見ることにより判断できる。メッセージの内容は対話規則に定義されているものでなければならないので、対話規則に規定されていないメッセージ及び対話の流れにあつていないメッセージを送信した場合には不正なエージェントである。また、ある期間内にシステムで許される数以上のメッセージを送信したかということも、メッセージを検査することにより検出できる。これによりメッセージの送信元が不正なエージェントであると判断された場合には、エージェントを非活動化する。例えば、エージェントを消滅させることも可能である。

【0013】以上をまとめると以下になる。すなわち、使用可能なメッセージを規定する対話規則を参照して、メッセージを検査し、当該検査結果に従って、メッセージの宛先エージェントを活動化候補エージェントとして指定する。例えば、検査結果が宛先エージェントを活動化すべきであるというものである場合に活動化候補エージェントとして指定される。このように対話規則を参照することによりエージェント・サーバにおける負荷を制御することができる。なお、メッセージが正常なものでエージェントの活動化が必要な時でも直ぐには活動化させない場合がある。これは、既に多数のエージェントが活動状態にある場合等のためである。すなわち、活動中のエージェントの状況を参照して、活動化候補エージェントを活動化する。

【0014】メッセージ検査の際には、使用可能なメッセージ及びエージェント間の対話を開始させるメッセージを定義する対話規則を参照して、メッセージが対話を開始させるためのものであるか検査する場合もある。メッセージがエージェント間の対話で使用可能なメッセージであっても、対話の流れの中で、対話を開始するためのメッセージでなければ、新たにエージェントを活動化させる必要はないからである。

【0015】また、活動化候補エージェントを指定する処理において、検査結果によりメッセージの宛先エージェントは活動化させるべきであると判断された場合には、活動化候補エージェントとして宛先エージェントの活動化を要求するステップと、活動化候補エージェントの活動化の可否を判断するステップとを含むようにすることも考えられる。

【0016】また、検査結果によりメッセージが不正なものであると判断された場合には、当該メッセージの送信元エージェントを非活動化することも考えられる。対話規則に従わないメッセージの送信元エージェントは不正なエージェントだからである。

【0017】また、第2エージェントから送信された複数のメッセージを検査し、当該複数のメッセージが所定の条件を満たしていない場合には、第2エージェントを非活動化することも可能である。第2エージェントが、

6

所定以上の数のメッセージを送信した場合等に対応できる。

【0018】検査結果によりメッセージの送信元エージェントを非活動化させるべきであると判断された場合には、送信元エージェントを非活動化候補エージェントに指定する処理をさらに含むようにすることも考えられる。これは、不正なエージェントだけでなく、対話規則から処理が終了の終了を示すメッセージを送信したエージェントを非活動化することにより、エージェント・サーバの負荷を減らすためである。

【0019】また、エージェントの可能な活動を規定したエージェント活動ポリシーをさらに記憶し、当該エージェント活動ポリシーを参照して、エージェントの活動化及び非活動化を実施することも考えられる。例えば、指定されたエージェントがメモリ及びスレッドを所定期間以上占有することができないようにエージェント活動ポリシーを設定し、その指定エージェントが所定期間以上に活動しているか否か判断することも可能である。さらに、あるエージェントを常に活動中にさせるようエージェント活動ポリシーを設定することも可能である。

【0020】さらに、他のコンピュータからのハイパーテキスト・トランスポート・プロトコル(HTTP)メッセージをエージェント用のメッセージに変換するHTTPゲートウェーを設け、当該HTTPメッセージを変換することにより生成されたメッセージを検査して、他のコンピュータに対応するエージェントのエージェント使用要求を出力するようにすることも可能である。エージェントとユーザがブラウザ(Browser)を用いて対話する場合に対応するためである。

【0021】以上本発明の処理のフローを説明したが、本発明はこれらの処理を実施する装置や、コンピュータにこれらの処理を実施させるプログラムの形態によっても実施可能である。このプログラムを、フロッピー・ディスクやCD-ROM等の記憶媒体又は他の形態の記憶装置に格納することができる。

【0022】

【発明の実施の形態】本実施例では、以下のような旅行情報を提供するシステムを例(図1)にして説明する。エージェント・サーバ1には、パッケージ・ツアー情報を提供する旅行代理店の店舗エージェント9a及び9bが活動している。利用者は、クライアント・コンピュータ5のWebブラウザ11を利用して、このエージェント・サーバ1にアクセスする。利用者が初めてアクセスするとその利用者用の顧客エージェント7aがエージェント・サーバ1に生成される。図1では、3つの顧客エージェント7a、7b及び7cが存在している。クライアント・コンピュータ5の利用者は、顧客エージェント7aをエージェント・サーバ1内に生成すると、この顧客エージェント7aに「10万円以内の航空会社XYZで行くホノルル4日間、2月14日出発のツアー」とい

10

20

30

40

50

う検索条件を入力し、検索を開始させる。図 1 の Web ブラウザ 1 1 には、上記の検索条件を入力した状態を示している。この検索条件はネットワーク 3 を介してエージェント・サーバ 1 の顧客エージェント 7 a に入力される。顧客エージェント 7 a は入力された検索条件をメッセージで各店舗エージェント 9 a 及び 9 b に通知する。

【 0 0 2 3 】 このメッセージに回答して、店舗エージェント 9 a は、「航空会社 X Y Z で行くホノルルの A B C ホテル滞在 4 日間、2 月 1 4 日出発、9. 9 万円」をメッセージで顧客エージェント 7 a に返す。また、店舗エージェント 9 b は、「航空会社 X Y Z で行くホノルル 4 日間ツアー、2 月 1 4 日出発、8 万円」をメッセージで送信元エージェント 7 a に返す。さらに、店舗エージェント 9 b は、「本日限り、航空会社 B C D で行くホノルル 4 日間、2 月 1 4 日出発、5 万円」というお勧め商品情報も顧客エージェント 7 a にメッセージで送信する。

【 0 0 2 4 】 顧客エージェント 7 a は、これらの情報をまとめて、検索結果として Web ブラウザ 1 1 に、検索結果メッセージが送信される。Web ブラウザ 1 1 は、検索結果メッセージの内容を利用者に提示する。利用者は、所望のツアーが見つかった場合には、申込みをすることもできるし、新たな条件を Web ブラウザ 1 1 に入力して、再度検索することを命じてよい。顧客エージェント 7 a を、利用者がエージェント・サーバ 1 にアクセスするごとに生成し、接続がなくなった時に消滅させることもできる。また、所定期間顧客エージェント 7 a を保存しておき、再度利用者がアクセスした時に、この顧客エージェント 7 a を再活動化させるようにすることも可能である。保存しておく際には、過去の検索結果も合わせて記憶しておき、再活動化させた時に Web ブラウザ 1 1 に過去の検索結果を送信することも可能である。

【 0 0 2 5 】 クライアント・コンピュータ 5 はこれまでに存在している通常のコンピュータであり、本発明は主にエージェント・サーバ 1 の構成である。図 2 にエージェント・サーバ 1 の構成例を示す。エージェント・サーバ 1 は、HTTP ゲートウエー 2 1、対話管理機構 2 3、対話管理機構 2 3 により参照される対話規則データベース (DB) 2 5、エージェント管理機構 2 7、エージェント管理機構 2 7 に参照されるエージェント活動ポリシー・データベース (DB) 2 9、二次記憶装置であるエージェント格納域 3 1 を含む。

【 0 0 2 6 】 HTTP ゲートウエー 2 1 は、クライアント・コンピュータ 5 の Web ブラウザ 1 1 と HTTP メッセージのやり取りを行う。図 3 に HTTP ゲートウエー 2 1 の動作を説明するための模式図を示す。Web ブラウザ 1 1 の入力欄に図 3 のように入力して送信を命ぜると、HTTP メッセージが生成されて出力される。例えば、<http://www.tabican.ne.jp:987/search/AID=1234/search.html?mnopqr> といったメッセージになる。こ

では、www.tabican.ne.jp 部分は、エージェント・サーバ 1 の IP (Internet Protocol) アドレスであり、987 部分は、その通信ポート番号を表す。また、AID=1234 は、エージェント・サーバ 1 における対応顧客エージェント番号であり、当該顧客エージェントにはメッセージ名 [search.html](#) で、mnopqr は入力データを表す。このような HTTP メッセージは、HTTP ゲートウエー 2 1 でエージェント用のメッセージに変換される。図 3 ではエージェントに直接渡されるようにされているが、これは HTTP ゲートウエー 2 1 の説明のためであって、図 2 に示したように、メッセージは対話管理機構 2 3 に渡される。一方、エージェントからのメッセージも利用者に出力する必要がある場合には、HTTP ゲートウエー 2 1 によって、HTML データを含む HTTP メッセージに変換される。そして、Web ブラウザ 1 1 にて利用者に提示される。なお、本実施例では、内容を利用者に出力する必要があるメッセージは、エージェント自身が HTML データを出力する。Web ブラウザ 1 1 を用いてエージェントと対話する技術については、IBM TDB Vol.40 No.08 pp127-129, 1997 年 8 月に記載されているように周知の技術である。

【 0 0 2 7 】 対話規則 DB 2 5 は、エージェント間で行われる対話規則の情報を保存するデータベースである。対話とは一連のメッセージ交換にて実現され、対話規則には使用されるメッセージのタイプ、引数、引数のタイプが定義され、一連のメッセージの順番も定義されている。例えば、図 4 の上側に定義の記述と、下側にそれらの意味を示す。図 4 の例では、各セッション (Session) ごとに、最も上がセッション開始のメッセージ、最も下がセッション終了のメッセージを示す。セッションとは、エージェント間で張られる対話のためにやり取りされる一連のメッセージ群である。HTTP ゲートウエー 2 1 も特殊なエージェントとして取り扱われるので、図 4 の左の一番上のセッションでは、Agent A が HTTP ゲートウエー 2 1 であり、Agent B がその他のエージェントである。

【 0 0 2 8 】 対話管理機構 2 3 は、対話規則 DB 2 5 を参照して、エージェント間の対話が正常なものであるかを判断する。すなわち、対話規則 DB 2 5 に定義されているメッセージのタイプ、引数、引数のタイプ、及びメッセージの順番である。この時、対話管理ブロックを用いて対話の状態遷移を保存しておく。これにより、現在何番目のメッセージまで送信済で次はどのメッセージがどのエージェントからどのエージェントへ送信されるかを判断する。そして、セッションの最初に定義された適正なメッセージを受信した場合には、宛先エージェントの使用要求を、エージェント管理機構 2 7 に出力する。そして、エージェント管理機構 2 7 からエージェント使用許可を受信した場合には、メッセージの宛先エージェントにメッセージを送信する。エージェントが活動中

10

20

30

40

50

で、適正なメッセージが送信されている場合には、エージェントへメッセージを送信する。一方、受信したメッセージがセッションの最後に定義された適正なメッセージを受信した場合には、送信元エージェントの使用終了通知をエージェント管理機構27に出力する。

【0029】図5及び図6並びに図9に、対話管理機構23の処理を詳細に示す。図5は、あるエージェントへのメッセージの送信依頼処理を表す。送信依頼とあるのは、どのメッセージも対話管理機構23を介してメッセージを送信するので、対話管理機構23ではメッセージの受信をメッセージ送信の依頼と解釈されるためである。最初に、対話規則における最初のメッセージであるかどうか判断する(ステップ103)。もし、対話規則における最初のメッセージである場合には、上述の対話管理ブロックを作成する(ステップ105)。そして、エージェント活動要求をメッセージ送信依頼に付加する(ステップ107)。エージェント活動要求は、メッセージの宛先エージェントに対する活動要求であり、この宛先エージェントは活動化候補エージェントとなる。そして、このメッセージ送信依頼を待ち行列に入れる(ステップ109)。一方、対話規則における最初のメッセージでない場合には、既にエージェントは活動化されており且つ対話管理ブロックは作成されているので、ステップ105及び107をスキップして、ステップ109に飛ぶ。そして、エージェント管理機構27にメッセージ送信依頼処理(後述)を実施させる(ステップ111)。

【0030】図6はエージェント管理機構27のメッセージ送信依頼処理においてメッセージを送信させる段階になったところで実施される、メッセージ送信処理を表す。メッセージ送信の最初の処理は、対話の流れ定義どおりのメッセージであるかどうかの判断である(ステップ123)。もし、対話の定義どおりのメッセージでない場合には、エラーとして処理される。例えば、メッセージの送信元エージェントを非活動化させる等の処置をとる。一方、メッセージが定義どおりであれば、実際にメッセージを送信し(ステップ127)、処理を終了する。

【0031】ここでは、メッセージが対話の流れ定義どおりであるかをステップ123にて判断しているが、図5におけるステップ103で合わせて判断することも可能である。

【0032】エージェント活動ポリシーDB29は、エージェントのグループごとに、常に活動状態にする、対話中のみ活動状態にする等のスケジュール・ポリシーを格納するものである。図10に例を示す。上側が定義の記述であり、下側がその定義の意味を表したものである。ここでは、Consumerエージェントは最大5つ同時に活動でき、Shopエージェントは2つである。Shop A エージェントは、常に活動し、それ以外は非活動化され

ることもある。また、Shop B エージェントは1回15秒しか活動できず、Consumerエージェントは30秒である。このエージェント活動ポリシーDB29は、エージェント管理機構27により参照される。

【0033】エージェント管理機構27は、エージェント活動ポリシーDB29を参照して、エージェントの活動化・非活動化のスケジュールと処理を実施する。すなわち、対話管理機構23からのエージェント使用要求

(図5におけるエージェント活動要求を含むメッセージ送信依頼)及びエージェント使用終了通知(後述:対話管理ブロックなしの判断)を受け取り、エージェント活動ポリシーDB29を参照して、適切なタイミングで活動化及び非活動化を実施する。特に、ポリシーDB29に活動するエージェントの数のみ記述されている場合には、規定の数以上のエージェントが活動化されないようにする。

【0034】対話管理機構23の処理を表した図5のステップ111の処理、すなわちエージェント管理機構27のメッセージ送信依頼の処理を図7に示す。まず、現在処理を行っているエージェントの数が制限値より小さいかどうか判断する(ステップ143)。この「処理を行っているエージェント」とは、実際にメッセージを受信して、そのメッセージに対する処理を実施しているエージェントである。対話管理機構23又はエージェント管理機構27は、処理を行っているエージェントの数を把握するために、エージェントごとに処理をしているメッセージの数をカウントするためのカウンタを管理している。例えば、エージェントが3つのメッセージの処理を実施している場合には、そのエージェントのカウンタは3となっている。このステップ143では、全てのエージェントのカウンタの値を加算した値を用いる。場合によっては、カウンタの値が1以上であるエージェントの数が用いられる。もし、ステップ143で制限値より小さい場合には、メッセージ送信依頼の待ち行列の先頭の依頼をチェックする(ステップ145)。そして、その先頭の依頼にエージェント活動要求が付加されているか判断する(ステップ147)。もし、エージェント活動要求が付加されている場合には、エージェントを活動化処理を実施する(ステップ149)。この活動化処理の詳細は図8を用いて説明する。なお、この図8を見れば直ぐに分かるが、エージェント活動化処理では、実際はエージェントが活動化されない場合がある。よって、このメッセージ送信依頼処理の中では、送信先エージェントは現在活動中であるかどうかを確認する(ステップ153)。活動化されていない場合には、メッセージの送信処理(図6)を実施できないので、エージェント管理機構27におけるメッセージ送信依頼処理を終了する(ステップ159)。

【0035】ステップ147においてエージェント活動要求が先頭の依頼に付加されていなかった場合及びステ

ップ153においてメッセージの送信先エージェントが現在活動中であると確認された場合には、メッセージ送信依頼待ち行列の先頭の依頼を取り出す(ステップ151)。そして、先ほど述べた、エージェントが処理しているメッセージの数(1以上であればエージェントが処理中であることを表している)を1インクリメントする(ステップ155)。そして、図6に示されたメッセージ送信処理を実施する(ステップ157)。そして、エージェント管理機構27のメッセージ送信依頼処理を終了する。

【0036】なお、ステップ143において、処理中のエージェントの数が制限値と同じ又はそれより大きい場合には、処理を終了する。この図7の処理は、後に述べる図9の処理から再度開始され、メッセージ送信依頼の待ち行列の先頭の依頼を次々に処理していく。

【0037】図8に図7のステップ149に出てきたエージェント活動化処理を示す。図7のステップ149に処理が移行すると、まず送信先エージェントは現在活動中か判断する(ステップ163)。もし、現在活動中であるならば、再度活動化する必要はないので、本処理を終了する(ステップ169)。一方、活動中でない場合、現在活動化されているエージェントの数が、制限値より小さいか判断する(ステップ165)。この活動化されているエージェントと処理中のエージェントの数は異なる。例えば、活動化されていても、メッセージの処理を現在行っていないエージェントも存在するからである。もし、制限値と同じ又はそれより大きい場合には、エージェントを活動化することなく処理を終了する(ステップ169)。一方、制限値より小さい場合に、エージェントを活動化して(ステップ167)、処理を終了する。

【0038】このように本発明では処理中のエージェントの数をも管理する。これは、システムにおける処理負荷は、エージェントの活動化だけではなく、実際にエージェントがメッセージの処理をしていることに起因するものも大きいからである。例えば、店舗エージェントは商品DBの検索を実行するため、二次記憶装置へアクセスする。この処理は他の店舗エージェントの処理と重なることも多い。よって、活動化エージェントの数のみならず、処理中のエージェントの数の管理も非常に重要である。

【0039】次に、メッセージ処理終了のための処理について図9を用いて説明する。この図9の一部は対話管理機構23が実施し、残りはエージェント管理機構27が実施するようになっている。なお、この処理の開始は、エージェントが1つのメッセージについての処理を終了して、その処理終了に対するメッセージを送信することによりスタートする。そして、この処理終了に対するメッセージが対話定義の最後のメッセージかを判断する(ステップ183)。もし、対話定義の最後のメッセ

ージである場合には、このメッセージに関する対話管理ブロックを削除する(ステップ185)。そして、送信元のエージェントの対話管理ブロックがまだ存在しているか判断する(ステップ187)。もし、送信元エージェントが他のメッセージの処理を実施している場合には、他のメッセージに関する対話管理ブロックが残っているので、ステップ191に移行する。一方、もう対話管理ブロックが存在しない場合には、エージェント管理機構27に通知して、エージェント管理機構27がエージェントの非活動化を実施する(ステップ189)。

【0040】なお、ステップ183において対話定義の最後のメッセージでないと判断された場合及びステップ187において対話管理ブロックがまだ存在すると判断された場合には、当該エージェントの処理中を表すカウンタを1デクリメントする(ステップ191)。これにより、システムにとっての負荷が軽くなったので、メッセージ送信依頼の待ち行列に未処理の依頼が存在するか判断する(ステップ193)。もし、存在しないならば、処理を終了する(ステップ197)。また、存在する場合には、待ち行列から1つ依頼を取り出し、図7のメッセージ送信依頼を実行する。

【0041】非活動化は、エージェント管理機構27がエージェント活動ポリシーDB29を参照して、独自に処理を開始する場合もある。同時に多数の対話開始を表すメッセージを受信した場合、エージェント管理機構27は、現在活動状態にあるエージェントを一時的に非活動化することにより、順次エージェントを活動状態にし、メッセージを送信することもある。これは、例えば全体で5つのエージェントが活動できる環境において、既に5つのエージェントが活動している場合に、あるエージェントが活動していないエージェントに対してメッセージを送信すると、もう新たにエージェントを活動化できない。よって、この場合には1つのエージェントを臨時に非活動化することにより処理を進める場合もある。また、例えば、図10のように1回15秒しか活動できない場合には、活動化の時にタイマをスタートさせ、15秒たったところで、非活動化処理を実施する。また、対話管理機構23が不正なメッセージを検出した場合には、当該メッセージの送信元エージェントを非活動化するように要求する。よって、これに回答して、送信元エージェントを非活動化してもよい。

【0042】実際にエージェントを活動化及び非活動化する処理は、オブジェクトのシリアライゼーション及びその逆を使用することができ、例えば「Pickling State in the Java System」(The 2nd USENIX Conference on Object-Oriented Technologies, 1996)に実装方法の一例が述べられている。

【0043】なお、図5乃至図9に示した処理例は、そ

10

20

30

40

50

13

れぞれ対話管理機構 2 3 又はエージェント管理機構 2 7 で処理の分担を示したが、両者が一体となって図 5 乃至図 9 の処理を実施するように変形可能である。また、対話管理機構 2 3 及びエージェント管理機構 2 7 を一体で実現することも可能であり、それにより図 5 乃至図 9 の処理を一体で実施することも可能である。

【0044】では図 1 で使用した例を図 1 1 及び 1 2 を用いて、全体の処理の流れを説明する。利用者は「10 万円以内で、航空会社 X Y Z でホノルル 4 日間、2 月 1 4 日出発のツアー」を検索するように HTTP メッセージ 2 0 1 をクライアント・コンピュータ 5 から出力する。この HTTP メッセージ 2 0 1 は、HTTP ゲートウエー 2 1 が HTTP Request メッセージ 2 0 3 に変換し、対話管理機構 2 3 に出力する。対話管理機構 2 3 は、対話規則 DB 2 5 を参照して、本メッセージがセッションの開始であると判断する。そうすると、対話管理機構 2 3 はエージェント使用要求（エージェント活動要求）2 0 5 をエージェント管理機構 2 7 に出力する。エージェント管理機構 2 7 は、活動化が可能なタイミングで Agent A を活動化（2 0 7）させ、対話管理機構 2 3 に使用許可（メッセージ送信処理への移行：図 7 のステップ 1 5 7）2 0 9 を出力する。使用許可がでると対話管理機構 2 3 は、HTTP Request 2 1 1 を Agent B に出力する。

【0045】この後は図 1 2 を用いて説明する。Agent A では、HTTP Request 2 1 1 内の検索条件を引数とする RequestTravelGoods メッセージ 2 2 3 を対話管理機構 2 3 に出力する。対話管理機構 2 3 は、対話規則 DB 2 5 を参照してこのメッセージ 2 2 3 が Agent B との対話の開始であることを判断し、Agent B の使用要求をエージェント管理機構 2 7 に出力する。そして、エージェント管理機構 2 7 は、Agent B を適切なタイミングで活動化させ、対話管理機構 2 3 に使用許可通知 2 2 9 を出力する。そうすると、対話管理機構 2 3 は、RequestTravelGoods メッセージ 2 3 1 を Agent B に出力する。Agent B は、このメッセージ 2 3 1 に対する処理を行い、その返事として「航空会社 X Y Z で行くホノルル 4 日間、2 月 1 4 日出発を 8 万円」という内容の ProvideExactGoods メッセージ 2 3 3 を、対話管理機構 2 3 に出力する。対話管理機構 2 3 は、対話規則 DB 2 5 を参照して、このメッセージ 2 3 3 が順番、引数等を含めて適正であるかを判断して、もし適正であれば Agent A へ ProvideExactGoods メッセージ 2 3 5 として出力する。さらに、Agent B は、「本日限り、航空会社 B C D で行くホノルル 4 日間、2 月 1 4 日出発、5 万円」というお勧め商品情報に関する RecommendGoods メッセージ 2 3 6 を対話管理機構 2 3 に出力する。対話管理機構 2 3 は、対話規則 DB 2 5 を参照すると、当該メッセージ 2 3 6 がセッションの終了を表すメッセージと判断する。そして、対話規則通りであれば Re

14

commendGoods メッセージ 2 3 7 を Agent A に出力し、エージェント B の使用終了（対話管理ブロックなし：図 9 ステップ 1 8 7 から 1 8 9）2 3 9 をエージェント管理機構 2 7 に通知する。エージェント管理機構 2 7 は、活動ポリシー DB 2 9 を参照して、Agent B が非活動化するエージェントであれば、Agent B を非活動化させる（2 4 1）。

【0046】もう一度図 1 1 に戻って、Agent A は必要な情報を得たので、利用者に検索結果を出力するための処理を実施する。そこで、HTTP Response メッセージ 2 1 3 を対話管理機構 2 3 に出力する。対話管理機構 2 3 は、対話規則 DB 2 5 を参照して、このメッセージ 2 1 3 が Agent A と HTTP ゲートウエー 2 1 とのセッションの最後のメッセージであると判断する。そうすると、対話管理機構 2 3 はエージェント管理機構 2 7 に Agent A の使用終了を通知する（2 1 7）。また、HTTP Response メッセージ 2 1 5 を HTTP ゲートウエー 2 1 に出力する。エージェント管理機構 2 7 は、活動ポリシー DB 2 9 を参照して、Agent A を非活動化できる場合には、Agent A を非活動化する（2 1 9）。HTTP ゲートウエー 2 1 は、対話管理機構 2 3 からの HTTP Response メッセージ 2 1 5 を HTTP メッセージ 2 2 1 に変換し、出力する。

【0047】これにより、クライアント・コンピュータ 5 には利用者が命じた検索結果が Web ブラウザ 1 1 に表示されるようになる。

【0048】以上実施例を説明したが、本発明の様々な変形が可能である。例えば、対話管理機構 2 3 が対話規則を自ら含むようにしたり、同じくエージェント管理機構 2 7 がエージェント活動ポリシーの規定を自ら含むようにすることもできる。先にも述べたように、対話管理機構 2 3 とエージェント管理機構 2 7 を一体のものとすることも可能である。さらに、HTTP ゲートウエー 2 1 は、HTTP が用いられていることを前提にしているが、他のプロトコルを用いる場合には、他のプロトコルとのゲートウエーとすることも可能である。また、HTTP ゲートウエー 2 1 を特別なエージェントとしていたが、対話規則 DB 2 5 においてエージェントでないように規定することも可能である。HTTP ゲートウエー 2 1 を複数設けることも可能である。さらに、エージェント記憶域 3 1 は、二次記憶装置でハードディスクを念頭にしているが、他の記憶装置であってもよい。但し、主記憶装置は、活動状態のエージェントのために使用しない方がよい。メッセージ送信依頼の待ち行列を用いるように説明したが、エージェント管理機構 2 7 において活動化待ちのエージェントの待ち行列を設けるようにすることも可能である。

【0049】不正なエージェントを検出した場合には、非活動化するのが好ましい。非活動化として、消滅させることも可能であるし、不正なエージェントであること

15

を表すタグを付けて正常なエージェントのようにエージェント格納域 31 に格納して、後に解析することも可能である。メッセージを複数検査したのちに、例えばメッセージ送信頻度が異常に高い等ということ判断して、不正なエージェントを判定することもできる。また、複数不正のメッセージがあった場合に不正なエージェントと判定するように、条件を緩和することもできる。また、対話規則は図 3 は一例であって、他の規則を定義することは自由である。対話を簡単にする場合には、RecommendGoods メッセージを定義しないようにすることは当然可能である。

【0050】

【効果】 エージェント・サーバの負荷を制御することができるようになった。

【0051】 メッセージを検査することにより、エージェントの活動／非活動を制御する方法を提供することができた。

【0052】 不正なエージェントを検出し、対応処理を実施することができた。

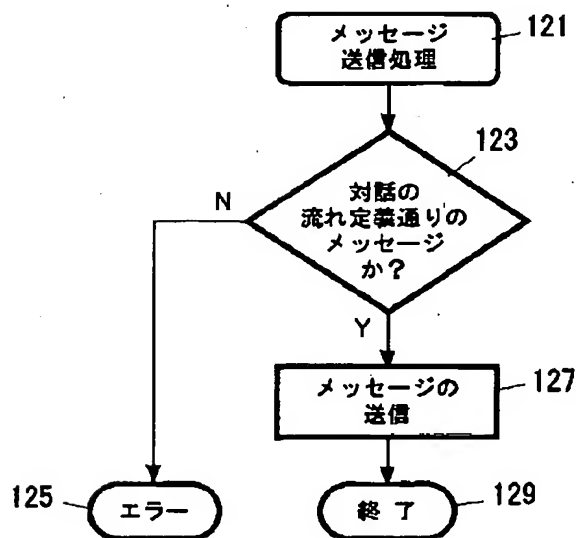
【0053】 メッセージを検査することにより、不正なメッセージを出力するエージェントを検出し、不正なエージェントを非活動化する方法を提供することができた。

【図面の簡単な説明】

【図 1】 本発明の概念図である。

【図 2】 本実施例におけるエージェント・サーバの構成例を示す図である。

【図 6】



16

【図 3】 HTTP ゲートウェイの処理を表す図である。

【図 4】 対話規則の例を示す図である。

【図 5】 対話管理機構によるメッセージ送信依頼の処理フローである。

【図 6】 メッセージ送信の処理フローである。

【図 7】 エージェント管理機構による、メッセージ送信依頼の処理フローである。

【図 8】 エージェント活動化の処理フローである。

【図 9】 メッセージ処理終了の処理フローである。

【図 10】 エージェント活動ポリシー DB の記述例である。

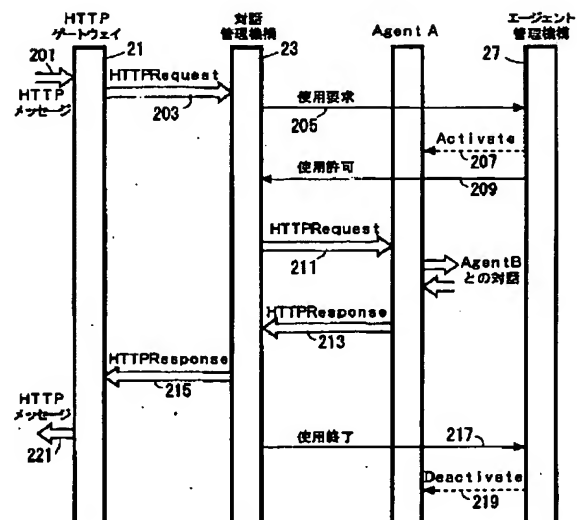
【図 11】 システム全体の処理の一例を表す図である。

【図 12】 システム全体の処理の一例を表す図である。

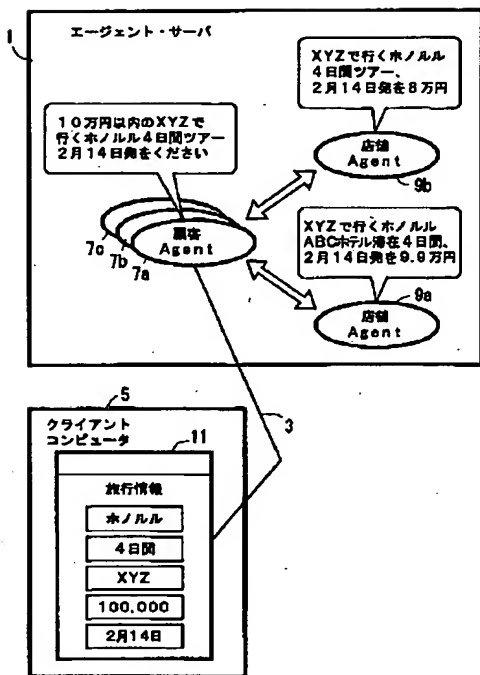
【符号の説明】

- 1 エージェント・サーバ
- 3 ネットワーク
- 5 クライアント・コンピュータ
- 7 顧客エージェント
- 9 店舗エージェント
- 11 Web ブラウザ
- 21 HTTP ゲートウェイ
- 23 対話管理機構
- 25 対話規則 DB
- 27 エージェント管理機構
- 29 エージェント活動ポリシー DB
- 31 エージェント格納域

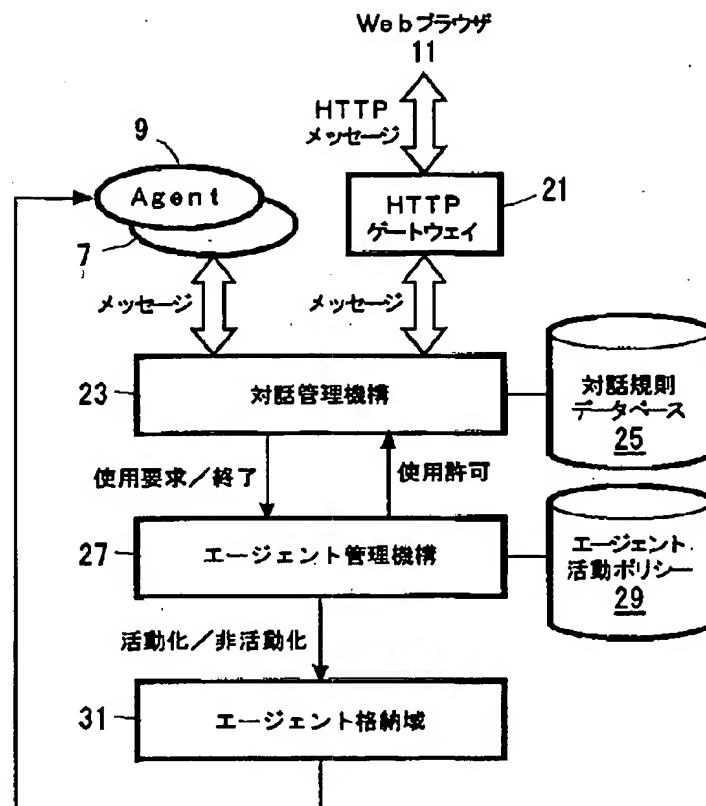
【図 11】



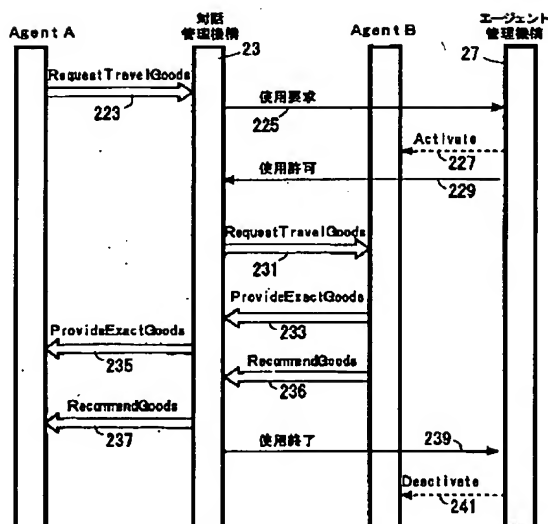
【図 1】



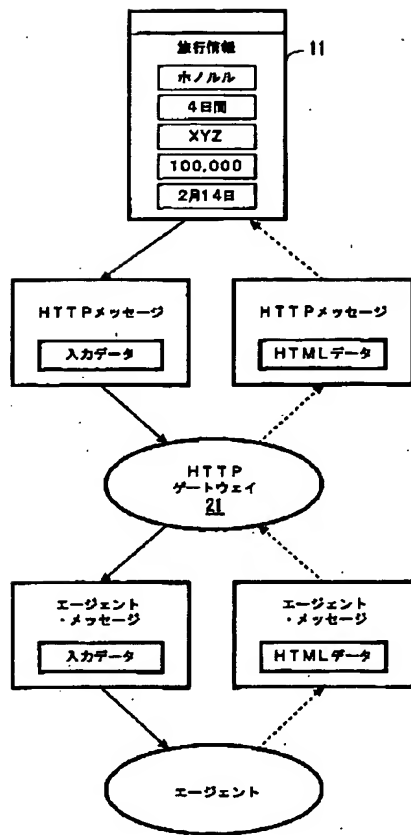
【図 2】



【図 1 2】



【図3】



【図4】

定義の記述

```

Define Session [from A to B]
  HTTPRequest [A to B]
  Method as String
  Parameters as Hashtable
  HTTPResponse [B to A]
  Contents as String

```

```

Define Session [from A to B]
  RequestTravelGoods [A, B]
  Requirement as QueryType
  ProvideExactGoods [B, A]
  Goods as Vector<Tour>
  RecommendGoods [B, A]
  Goods as Vector<Tour>

```

```

Define Type
  Category Tour
  Attribute AgencyID as String
  Attribute ProductID as String
  Attribute ProductName as String

```

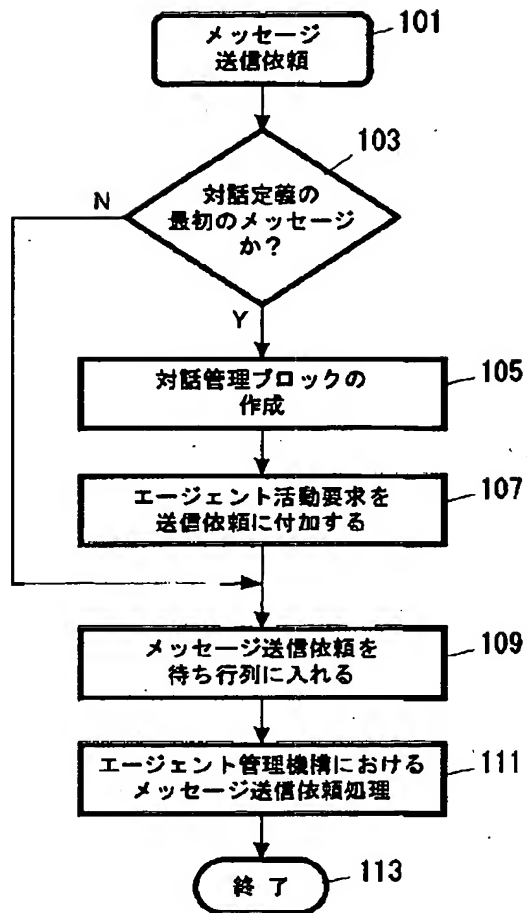
定義の意味

セッションの定義 [Agent AからAgent Bへ]
 ブラウザからの要求 [Agent AからAgent Bへ]
 引数名は Method でタイプは String
 引数名は Parameter でタイプは Hashtable
 ブラウザへの応答 [Agent BからAgent Aへ]
 引数名は Contents でタイプは String

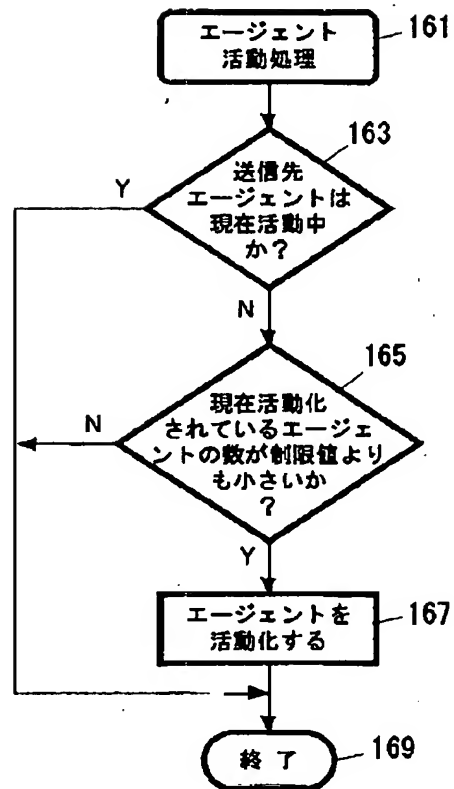
セッションの定義 [Agent AからAgent Bへ]
 旅行商品の要求 [AからBへのメッセージ]
 引数名は Requirement でタイプは QueryType
 条件に一致した旅行商品の提供 [BからAへのメッセージ]
 引数名は Goods でタイプは要素が Tour のベクタ
 お勧め旅行商品の提供 [BからAへのメッセージ]
 引数名は Goods でタイプは要素が Tour のベクタ

タイプの定義
 タイプ名は Tour
 属性名 AgencyID、タイプは String
 属性名 ProductID タイプは String
 属性名 ProductName タイプは String

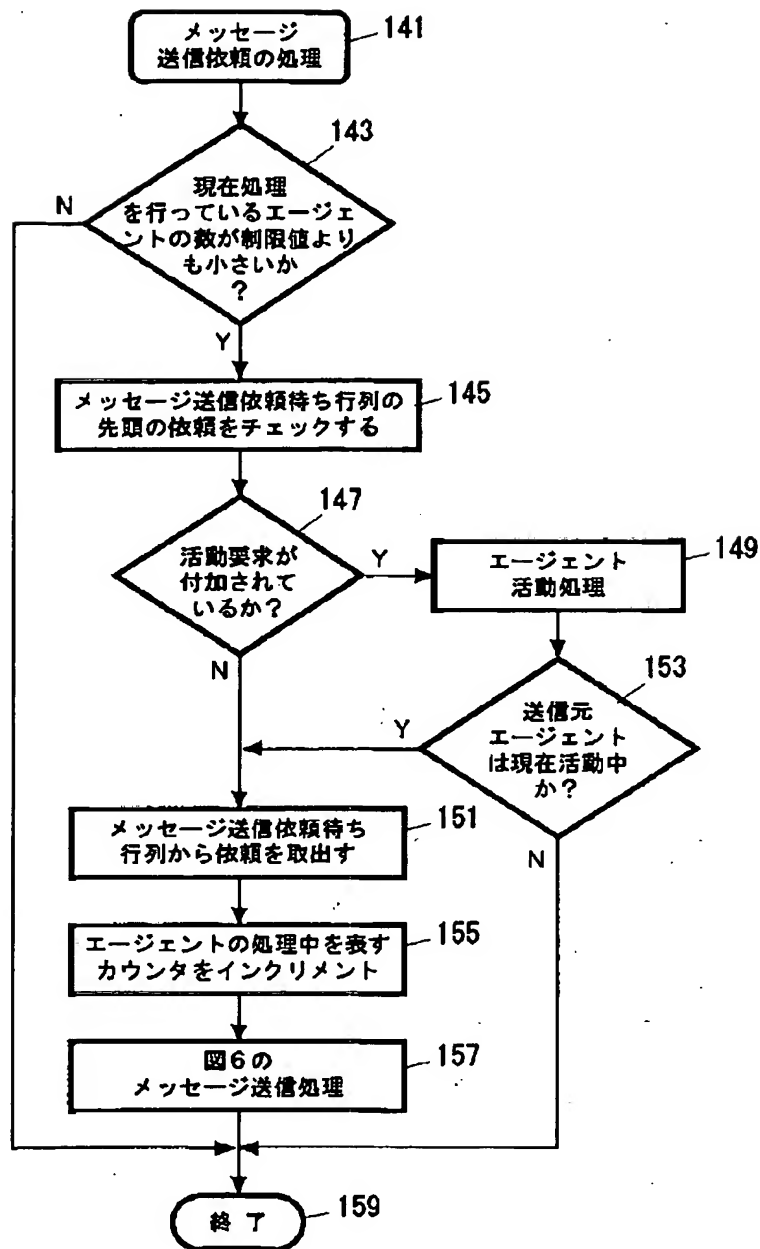
【図 5】



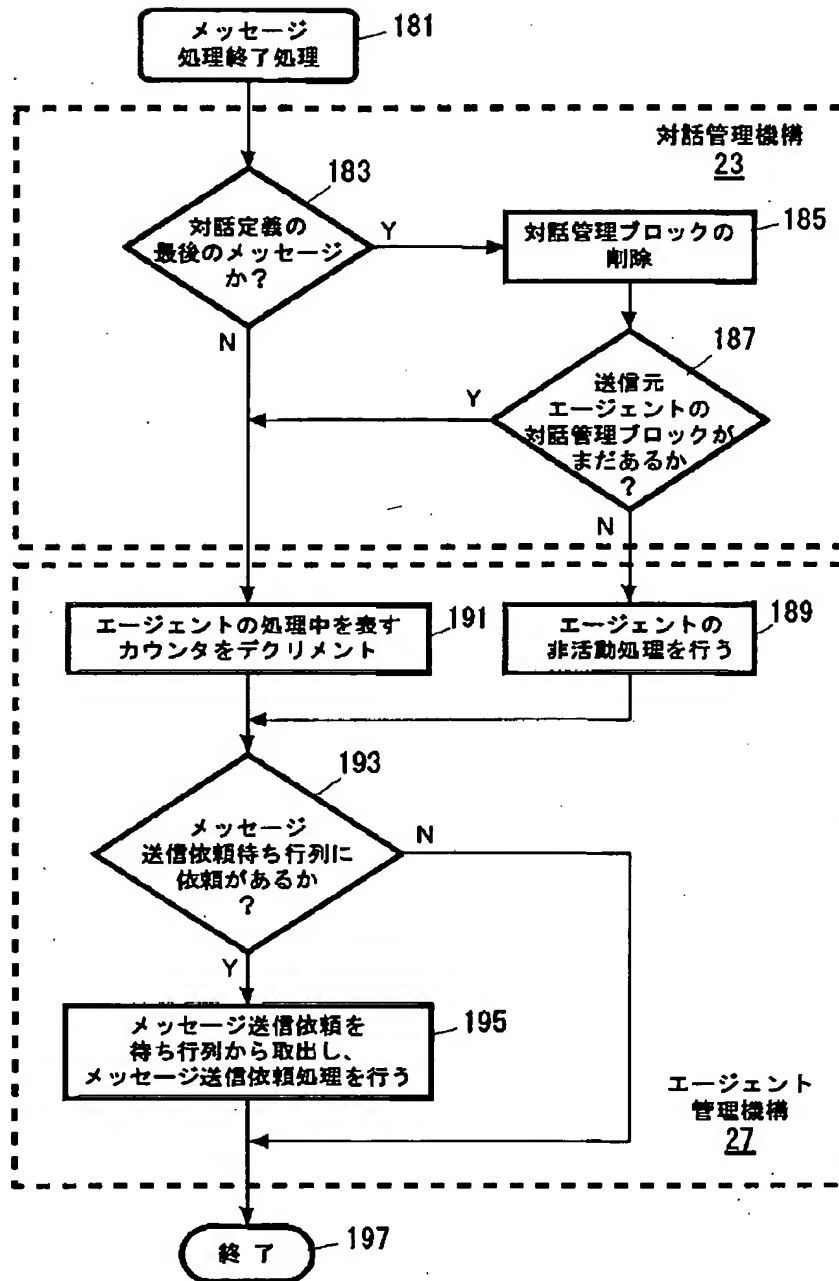
【図 8】



【図 7】



【図 9】



【図 1 0】

定義の記述

```
Define Group
  Name Consumer
  MaxActiveAgents 5
Define Group
  Name Shop
  MaxActiveAgent 2
Define Activity
  Name ShopA
  Group Shop
  ActivePolicy onMemory
Define Activity
  Name ShopB
  Group Shop
  ActivePolicy onDisk
  Timeout 15 sec
Define Activity
  Class com. ibm. eplace. Consumer
  Group Consumer
  ActivePolicy onDisk
  Timeout 30 sec
```

定義の意味

グループの定義
グループ名は Consumer
最大同時稼働数は 5

グループの定義
グループ名は Shop
最大同時稼働数は 2

活動方法の定義
対象 Agent の名前 は ShopA
グループは Shop
常に活動

活動方法の定義
対象 Agent の名前 は ShopB
グループは Shop
活動 / 非活動化を行う
1 回の活動制限時間は 15 秒

活動方法の定義
対象 Agent のクラス名は com. ibm. eplace. Consumer
グループは Consumer
活動方法は活動 / 非活動化を行う
1 回の活動制限時間は 30 秒

【手続補正書】

【提出日】平成 1 1 年 1 月 1 9 日

【手続補正 1】

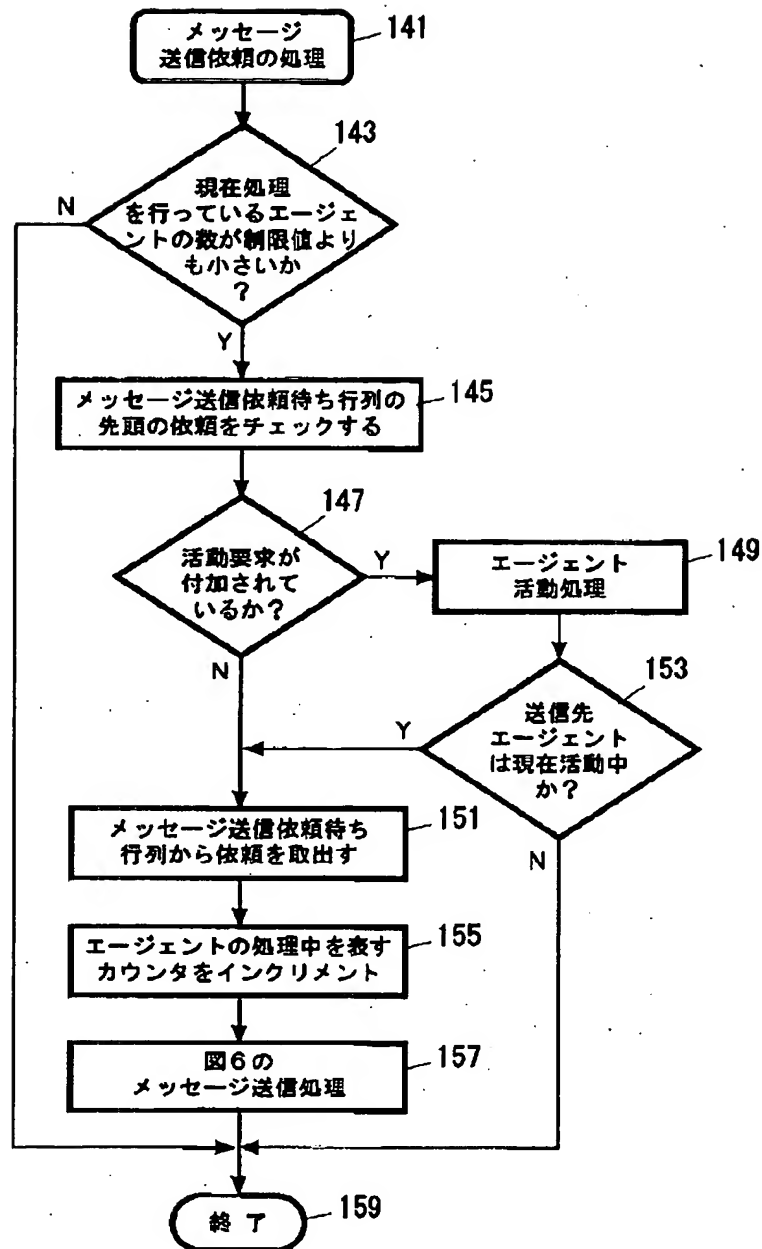
【補正対象書類名】図面

【補正対象項目名】図 7

【補正方法】変更

【補正内容】

【図 7】



フロントページの続き

(72)発明者 中村 祐一

神奈川県大和市下鶴間1623番地14 日本ア
イ・ビー・エム株式会社 東京基礎研究所
内